



# Temporal changes in field calibration relationships for Aeroqual S500 O<sub>3</sub> and NO<sub>2</sub> sensor-based monitors

Nicola Masey<sup>a</sup>, Jonathan Gillespie<sup>a</sup>, Eliani Ezani<sup>a,b</sup>, Chun Lin<sup>c</sup>, Hao Wu<sup>c</sup>, Neil S. Ferguson<sup>a</sup>, Scott Hamilton<sup>d</sup>, Mathew R. Heal<sup>c</sup>, Iain J. Beverland<sup>a,\*</sup>

<sup>a</sup> Department of Civil and Environmental Engineering, University of Strathclyde, James Weir Building, 75 Montrose Street, Glasgow, G1 1XJ, UK

<sup>b</sup> Department of Environmental and Occupational Health, Faculty of Medicine and Health Science, Universiti Putra Malaysia, 43400, Serdang, Selangor, Malaysia

<sup>c</sup> School of Chemistry, Joseph Black Building, University of Edinburgh, David Brewster Road, Edinburgh, EH9 3FJ, UK

<sup>d</sup> Ricardo Energy and Environment, 18 Blythswood Square, Glasgow, G2 4BG, UK

## ARTICLE INFO

### Keywords:

Air pollution  
Sensors  
Aeroqual  
Nitrogen dioxide  
Ozone

## ABSTRACT

Sensor-based monitors are increasingly used to measure air pollutant concentrations, but require calibration under field conditions. We made intermittent comparisons (6 times over a 6-month period) between ozone and nitrogen dioxide concentrations measured by Aeroqual gas-sensitive semiconductor (O<sub>3</sub>) and electrochemical (NO<sub>2</sub>) sensors (two of each) and reference analysers in the UK Automatic Urban and Rural Network. Each deployment period was split into equal ( $n = 48 \times 1$ -hour) training and test datasets, to derive and test calibration equations respectively. We observed significant bivariate linear relationships between *Aeroqual* O<sub>3</sub> and *Reference* O<sub>3</sub> concentrations, and significant multiple linear relationships between *Aeroqual* NO<sub>2</sub> and both *Reference* NO<sub>2</sub> and *Aeroqual* O<sub>3</sub> concentrations. Changes in monitor responses over time (including apparent baseline drift in O<sub>3</sub> sensor output, and discrepancies between the 2 Aeroqual NO<sub>2</sub> sensors) resulted in relatively inaccurate concentrations estimates (*cf.* reference concentrations) from calibration equations derived in the first training period and applied to subsequent *test* deployments (*e.g.* NO<sub>2</sub> RMSE =  $47.2 \mu\text{g m}^{-3}$  ( $n = 286$ ) for a dataset of all *test* periods combined, for one of the two monitor pairs). Substantial improvements in accuracy of estimated concentrations were achieved by combination of repeated intermittent training data into a single calibration dataset (NO<sub>2</sub> RMSE =  $8.5 \mu\text{g m}^{-3}$  for same *test* dataset described above). This latter approach to field calibration is recommended.

## 1. Introduction

The concentrations of gaseous air pollutants with deleterious health effects, including nitrogen dioxide (NO<sub>2</sub>) and ozone (O<sub>3</sub>), are monitored to estimate human exposure and to assess compliance with legislation and guidelines. Monitoring is usually conducted at static, automatic monitoring stations that record concentrations at high temporal resolution. However, the cost of these stations, together with practical considerations regarding suitable sites, invariably limits the spatial coverage of automatic networks. Wider geographical networks of passive diffusion samplers (PDS) can mitigate restrictions on spatial coverage [1,2]; however PDS provide limited temporal information, and can be subject to measurement inaccuracies associated with changing meteorological [3] and atmospheric chemistry conditions [4].

Battery-powered real-time hand-held sensor-based monitors for air

pollutants are continually being developed, which have potential to supplement data from existing monitoring networks [5]. Hand-held monitors usually have lower capital costs than automatic analysers, meaning that more could be made available for deployment, potentially increasing the spatial resolution of measurement networks with high temporal resolution concentration measurements [6–9]. The use of such instruments in mobile and personal monitoring has been reported [8,10].

The reference analysers in automatic monitoring stations are usually subject to documented quality control and assurance (QC/QA) procedures that aim to ensure the recorded pollutant concentrations are within specified ranges of accuracy and precision. The control of uncertainty requires resource and effort, so it is not unreasonable to anticipate that outputs from lower-cost, portable monitors are also subject to uncertainties, which may also vary with time. Thus, it is important that portable monitors are subject to

\* Corresponding author.

E-mail address: [iain.beverland@strath.ac.uk](mailto:iain.beverland@strath.ac.uk) (I.J. Beverland).

<https://doi.org/10.1016/j.snb.2018.07.087>

Received 27 June 2017; Received in revised form 10 May 2018; Accepted 16 July 2018

Available online 18 July 2018

0925-4005/ © 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

calibration checks at least as much as reference analysers. An important additional potential limitation of gas sensor-based monitors is their cross-sensitivity to other pollutants and/or to changing environmental conditions [11,12]. For example, the responses of some electrochemical sensors have been shown to be susceptible to variations in temperature or relative humidity [8,13–15]. Sensors for NO<sub>2</sub> have been shown to be cross-sensitive to O<sub>3</sub>, meaning that both pollutants must be measured simultaneously to allow correction of the NO<sub>2</sub> sensor response [8–10,16–18]. Consequently, for accurate estimation of ambient concentrations, sensor-based monitors require field calibration under representative ambient conditions.

In this study we evaluated the responses of two pairs of Aeroqual S500 O<sub>3</sub> and NO<sub>2</sub> monitors over time and repeated exposure to outdoor conditions, by deployment adjacent to reference gas analysers. The monitor deployments were designed to be representative of their likely use in practical field measurements: namely repeated cycles of calibration and monitoring (involving switching monitors on and off; removal from a reference site and use in other field measurements at different locations and mobile monitoring; and return to the reference monitoring site). Additionally, we investigated the effects of combination and timing of field calibration of the monitors on the accuracy of calibrated estimates. Our analyses provide insight to the refinement and use of field calibration relationships for these Aeroqual monitors in mobile measurements across geographical areas with widely varying pollution microclimates.

## 2. Methods

We evaluated two pairs of Aeroqual ([www.aeroqual.com](http://www.aeroqual.com)) S500 O<sub>3</sub> and NO<sub>2</sub> monitors at the UK Automatic Urban and Rural Network (AURN) Townhead urban background site in Glasgow city centre (55.866°N, 4.244°W; Location map given at: [http://www.scottishairquality.co.uk/latest/site-info?site\\_id=GLKP](http://www.scottishairquality.co.uk/latest/site-info?site_id=GLKP)). Although this site is in an urban background location it is within 1 km of a very busy motorway and intermittently experiences pollution episodes making it well-suited for calibration of sensors. Hourly-average concentration data from the reference analysers (API200A chemiluminescence analyser for NO<sub>2</sub> and Thermo 49i photometric analyser for O<sub>3</sub>) at this site were downloaded from <http://www.scottishairquality.co.uk/>. All AURN measurements were subject to documented national QC/QA procedures [19].

The Aeroqual monitors were delivered to our laboratory in June 2014. We deployed the monitors within ventilated waterproof enclosures provided by Aeroqual by attaching to the galvanised steel safety railings surrounding the roof of the AURN monitoring cabin (Fig. A1 in Appendix A: Supplementary Information). The NO<sub>2</sub> monitors contained electrochemical sensors (ENW2, range 0–1 ppm) and are referred to as NO<sub>2.1</sub> and NO<sub>2.2</sub>. The O<sub>3</sub> monitors contained gas-sensitive semiconductor sensors (OZU2, range 0–0.15 ppm) and are referred to as O<sub>3.3</sub> and O<sub>3.4</sub>. Monitors NO<sub>2.1</sub> and O<sub>3.3</sub> were located next to one another on the eastern railing while monitors NO<sub>2.2</sub> and O<sub>3.4</sub> were located on the western railing. Mains power was available allowing the monitors to operate continuously. We set the monitors to record gravimetric concentrations (µg/m<sup>3</sup>) at 1-min intervals, prior to computation of hourly-average concentrations for comparison with the analysers.

Six separate monitor co-location deployment periods were undertaken intermittently over 6 months (November 2015–May 2016) (Table A1 in Supplementary material). We truncated each deployment period to the first 96 h of field deployment to simplify comparison between periods.

## 3. Results

### 3.1. Calibration of Aeroqual O<sub>3</sub> monitors

Aeroqual O<sub>3</sub> measurements closely followed temporal trends in the Reference O<sub>3</sub> concentrations, with O<sub>3.3</sub> generally over-predicting and O<sub>3.4</sub> generally under-predicting the analyser concentrations (Fig. 1).

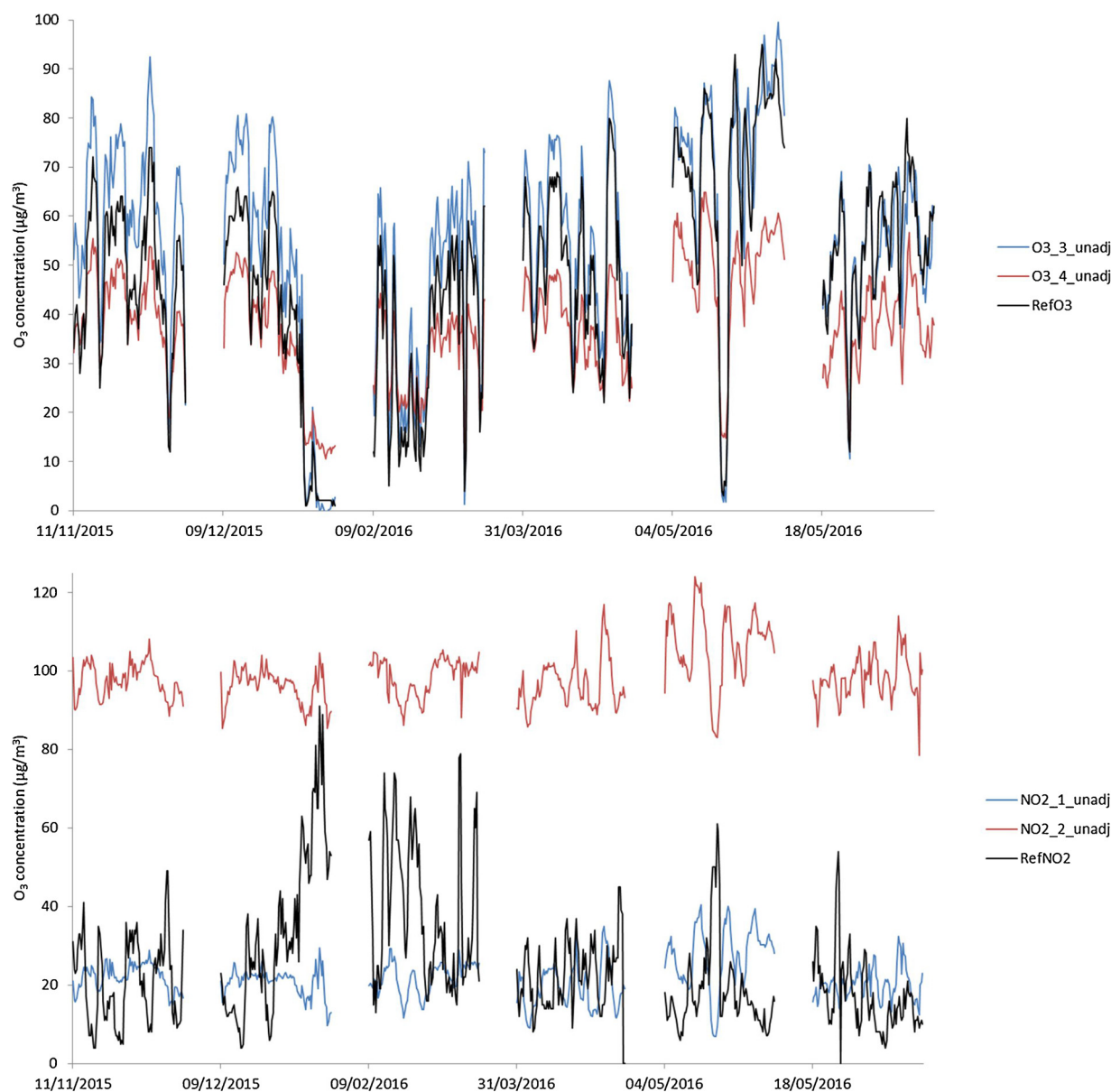
We calibrated the response of the O<sub>3</sub> monitors by calculating Ordinary Least Squares (OLS) regression equations between unadjusted Aeroqual O<sub>3</sub> measurements and O<sub>3</sub> concentrations measured by the reference analyser. We compared three methods of calibration - in each method the first 48 h for each deployment period was used as 'training' data to generate calibration equations, and the second 48 h was used as 'test' data to evaluate the accuracy of the calibrated predictions.

- 1) The first calibration method used the training data [0–48 h] for each deployment period to correct the test data [48–96 h] for the same deployment period *i.e.* a *unique* calibration for each deployment period (referred to as '*Aq\_corr\_u*'). This method represented calibrations computed at close time intervals to the measurements being corrected.
- 2) The second calibration method used a combination of training data from *all* deployment periods to derive a single global calibration equation for each monitor, which was then applied to the training data for each period (referred to as '*Aq\_corr\_a*'). This method of calibration, involving interspersed intervals during the extended set of deployments, had a larger number of data points in the calibration, spread over a longer time period, and thus incorporated a larger range of pollution conditions.
- 3) The third calibration method used the calibration equation derived from the training data from the first study period (*November*) to correct the test data from all of the subsequent studies (referred to as '*Aq\_corr\_N*'). This method represented a 'one-off pre-measurement campaign' calibration, and was included to assess how the accuracy and precision of a single calibration might deteriorate over time during extended field measurements.

The calibration equations are summarised in Table 1 and are shown in Fig. A3 in Supplementary material. Similar to the findings observed by Lin et al. [16] we observed limited and inconsistent effects of temperature and relative humidity on the calibration regression equations across the monitors and deployments and therefore these meteorological variables were not included in our analyses.

The calibration equations calculated for each of the three methods were used to adjust the test data. We observed similar overall temporal patterns in the time series of adjusted Aeroqual and Reference O<sub>3</sub> concentrations (adjusted concentration estimates shown for both *training* and *test* data in Fig. A4 in Supplementary material). The Reference O<sub>3</sub> and adjusted Aeroqual O<sub>3</sub> concentrations deviated for a short period at the end of the December deployment, when the reference analyser measured very low O<sub>3</sub> concentrations and the adjusted Aeroqual values were slightly negative. This period corresponded to a winter pollution episode with elevated NO<sub>2</sub> concentrations (Fig. A5 in Supplementary material). None of the three calibration methods accurately adjusted the Aeroqual measurements at these very low O<sub>3</sub> concentrations, presumably as a result of the marked change in pollution concentrations between the training and test data.

All three calibration methods yielded high correlation coefficients ( $R^2 > 0.90$ ) and slopes close to 1 (slope: 0.91–1.07) for the evaluation of adjusted test data against analyser-measured O<sub>3</sub> concentrations (Figs. 2 and S7). However, the scatter plot for the *Aq\_corr\_N* calibration (for both Aeroqual units) had a relatively large negative intercept compared to the scatter plots for the other two calibration methods,



**Fig. 1.** (a) Time series of unadjusted hourly-averaged O<sub>3</sub> concentrations measured by the two Aeroqual O<sub>3</sub> monitors and by the reference analyser for the full duration of each deployment period. Different deployment periods are separated by gaps in time series. The coefficient of determination ( $R^2$ ) between the concentrations measured by duplicate Aeroqual O<sub>3</sub> instruments was 0.90. (b) Time series of unadjusted hourly-averaged NO<sub>2</sub> concentrations measured by the two Aeroqual NO<sub>2</sub> monitors and by the reference analyser for the full duration of each deployment period. The coefficient of determination ( $R^2$ ) between the concentrations measured by duplicate Aeroqual NO<sub>2</sub> instruments was 0.83.

resulting from underestimation of reference analyser O<sub>3</sub>. The root mean square error (RMSE), mean bias (MB) and normalised mean bias (NMB) for *Aq\_corr\_N* were correspondingly larger than for the other two calibration methods (Table 2, Fig. A7 in Supplementary material).

These evaluation statistics indicated that calibrations derived from training data interspersed regularly throughout the deployment periods provided a more accurate estimate analyser concentrations in the test periods than a single calibration at the start of the deployment periods. For these O<sub>3</sub> monitors there were only small differences in  $R^2$ , RMSE, MB, and NMB statistics between *Aq\_corr\_u* and *Aq\_corr\_a* methods, with a slightly better fit between adjusted sensor and analyser data for the *Aq\_corr\_u* method.

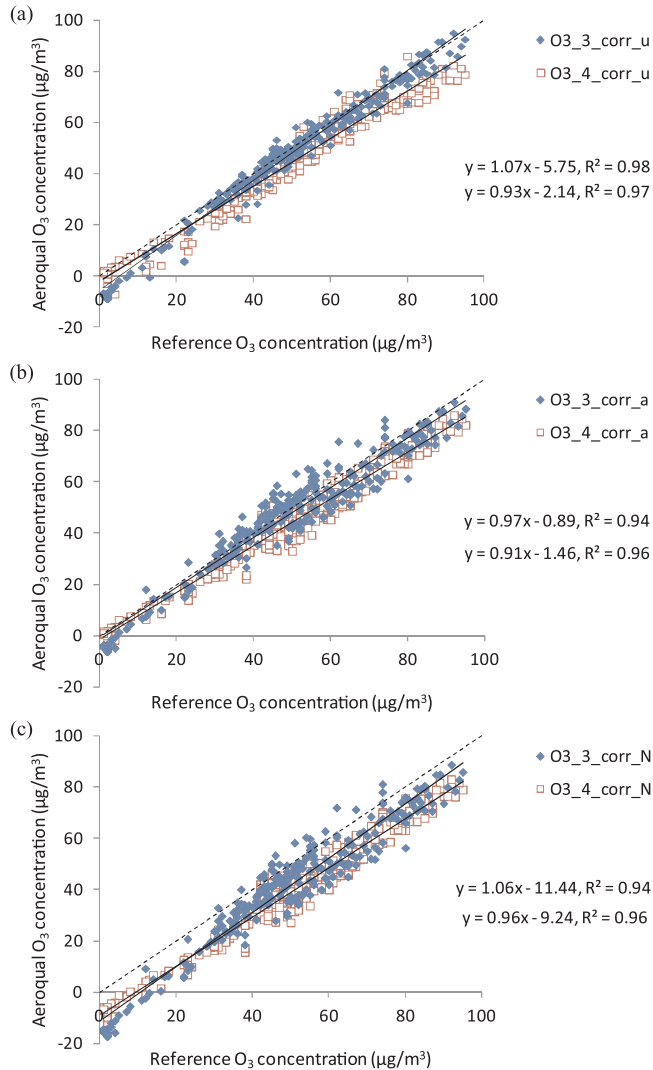
### 3.2. Calibration of Aeroqual NO<sub>2</sub> monitors

We established in our previous research [16] that the Aeroqual gas-sensitive semiconductor O<sub>3</sub> sensor responds linearly to ambient O<sub>3</sub> concentrations; and that the Aeroqual electrochemical sensor for NO<sub>2</sub> responds to both ambient NO<sub>2</sub> and ambient O<sub>3</sub>. For both gases, and both types of sensors, the sensor responses are based on conductivity/current changes (albeit changes of a somewhat different nature for the different sensor types) calibrated for exposure to different concentrations of individual gases by the manufacturer. This provides a physical basis for investigation of responses to both ambient O<sub>3</sub> and NO<sub>2</sub> as potential determinants of conductivity/current change calibrated to monitor outputs. We previously showed that the Aeroqual NO<sub>2</sub> monitor

**Table 1**

OLS equations for calibration equations for each training period (equivalent to *aq\_corr\_u*) and a global calibration for a combination of ‘All’ training periods (*aq\_corr\_a*) for the two pairs of Aeroqual O<sub>3</sub> and NO<sub>2</sub> monitors.

Pollutant	Study period	O <sub>3</sub> .3	O <sub>3</sub> .4
O <sub>3</sub>	11–15 Nov 2015	$0.94 * \text{Ref}_{\text{O}_3} + 16.11, R^2 = 0.85$	$0.53 * \text{Ref}_{\text{O}_3} + 16.73, R^2 = 0.95$
	9–13 Dec 2015	$1.07 * \text{Ref}_{\text{O}_3} + 9.34, R^2 = 0.96$	$0.59 * \text{Ref}_{\text{O}_3} + 12.19, R^2 = 0.90$
	9–15 Feb 2016	$1.08 * \text{Ref}_{\text{O}_3} + 5.41, R^2 = 0.98$	$0.51 * \text{Ref}_{\text{O}_3} + 15.27, R^2 = 0.96$
	31 Mar–4 Apr 16	$1.06 * \text{Ref}_{\text{O}_3} + 3.61, R^2 = 0.99$	$0.53 * \text{Ref}_{\text{O}_3} + 13.12, R^2 = 0.94$
	4–8 May 2016	$1.05 * \text{Ref}_{\text{O}_3} - 0.07, R^2 = 0.99$	$0.60 * \text{Ref}_{\text{O}_3} + 11.20, R^2 = 0.98$
	18–22 May 2016	$1.00 * \text{Ref}_{\text{O}_3} + 0.83, R^2 = 0.99$	$0.60 * \text{Ref}_{\text{O}_3} + 5.41, R^2 = 0.95$
	Combined	$1.02 * \text{Ref}_{\text{O}_3} + 6.26, R^2 = 0.93$	$0.56 * \text{Ref}_{\text{O}_3} + 12.66, R^2 = 0.91$
Pollutant	Study period	NO <sub>2</sub> .1	NO <sub>2</sub> .2
NO <sub>2</sub>	11–15 Nov 2015	$0.10 * \text{Ref}_{\text{NO}_2} + 0.17 * \text{O}_{3.3} + 8.99, R^2 = 0.34$	$0.38 * \text{Ref}_{\text{NO}_2} + 0.84 * \text{O}_{3.4} + 53.78, R^2 = 0.50$
	9–13 Dec 2015	$0.32 * \text{Ref}_{\text{NO}_2} + 0.28 * \text{O}_{3.3} - 3.29, R^2 = 0.33$	$0.07 * \text{Ref}_{\text{NO}_2} + 0.19 * \text{O}_{3.4} + 87.99, R^2 = 0.02$
	9–15 Feb 2016	$0.51 * \text{Ref}_{\text{NO}_2} + 0.49 * \text{O}_{3.3} - 18.95, R^2 = 0.92$	$0.35 * \text{Ref}_{\text{NO}_2} + 1.02 * \text{O}_{3.4} + 51.98, R^2 = 0.63$
	31 Mar–4 Apr 16	$0.50 * \text{Ref}_{\text{NO}_2} + 0.48 * \text{O}_{3.3} - 20.24, R^2 = 0.72$	$0.15 * \text{Ref}_{\text{NO}_2} + 0.53 * \text{O}_{3.4} + 70.82, R^2 = 0.36$
	4–8 May 2016	$0.54 * \text{Ref}_{\text{NO}_2} + 0.57 * \text{O}_{3.3} - 23.28, R^2 = 0.91$	$0.38 * \text{Ref}_{\text{NO}_2} + 1.05 * \text{O}_{3.4} + 47.73, R^2 = 0.93$
	18–22 May 2016	$0.17 * \text{Ref}_{\text{NO}_2} + 0.27 * \text{O}_{3.3} + 2.32, R^2 = 0.63$	$0.15 * \text{Ref}_{\text{NO}_2} + 0.66 * \text{O}_{3.4} + 71.21, R^2 = 0.74$
	Combined	$0.42 * \text{Ref}_{\text{NO}_2} + 0.43 * \text{O}_{3.3} - 13.19, R^2 = 0.67$	$0.33 * \text{Ref}_{\text{NO}_2} + 0.81 * \text{O}_{3.4} + 58.00, R^2 = 0.62$



**Fig. 2.** Scatter plots of adjusted hourly Aeroqual O<sub>3</sub> concentration estimates vs. reference analyser O<sub>3</sub> concentrations for the test data (*i.e.* 2nd half of deployment periods only). (a) *O<sub>3</sub>aq* adjusted using calibration derived from training data in each unique deployment period (*O<sub>3</sub>aq\_corr\_u*). (b) *O<sub>3</sub>aq* adjusted using calibration derived from training data combined from all deployment periods (*O<sub>3</sub>aq\_corr\_a*). (c) *O<sub>3</sub>aq* adjusted using calibration derived from the first deployment period (November) only (*O<sub>3</sub>aq\_corr\_N*).

sensitivity to O<sub>3</sub> concentrations could be characterised by a linear relationship between (*Aeroqual*NO<sub>2</sub> – *Reference*NO<sub>2</sub>) vs. *Aeroqual*O<sub>3</sub> which we used to correct the Aeroqual NO<sub>2</sub> concentrations. [16] Preliminary investigations of the Aeroqual NO<sub>2</sub> monitors used in the present work also revealed clear sensitivity of output on O<sub>3</sub> concentration. In our previous method [16], the coefficient for *Reference*NO<sub>2</sub> was effectively constrained to 1, which reflects the expectation that when the Aeroqual sensors are new and recently factory-calibrated their output concentration values should be close to True (*i.e.* reference) NO<sub>2</sub> concentrations:

$$\text{Aeroqual NO}_2 - \text{True NO}_2 = a_1 * \text{Aeroqual O}_3 + b_1 \quad (1)$$

Therefore, we can calibrate Aeroqual NO<sub>2</sub> by rearranging Eq. (1):

$$\text{True NO}_2 = \text{Aeroqual NO}_2 - a_1 * \text{Aeroqual O}_3 - b_1 \quad (2)$$

A potential shortcoming of this approach is that it does not allow for an Aeroqual NO<sub>2</sub> monitor to have a relationship to True NO<sub>2</sub> (*i.e.* to reference analyser NO<sub>2</sub>) that is not equal to 1; for example if the response of the sensor in the NO<sub>2</sub> monitor has reduced in time since its factory calibration. A more general model for the Aeroqual NO<sub>2</sub> monitor response is therefore based on the following two underlying relationships:

- 1) Aeroqual O<sub>3</sub> has a linear response to True (reference analyser) O<sub>3</sub>, *i.e.*

$$\text{Aeroqual O}_3 = m_1 * \text{True O}_3 + c_1 \quad (3)$$

- 1) Aeroqual NO<sub>2</sub> has linear responses to both NO<sub>2</sub> and O<sub>3</sub>, and with its response to O<sub>3</sub> being different to the response of the Aeroqual O<sub>3</sub> monitor to O<sub>3</sub>, *i.e.*

$$\text{Aeroqual NO}_2 = m_2 * \text{True NO}_2 + m_3 * \text{True O}_3 + c_2 \quad (4)$$

Substituting for True O<sub>3</sub> from Eq. (3) into Eq. (4) yields:

$$\text{Aeroqual NO}_2 = m_2 * \text{True NO}_2 + \frac{m_3}{m_1} (\text{Aeroqual O}_3 - c_1) + c_2 \quad (5)$$

Rearrangement of Eq. (5) into a similar form to Eq. (2) gives:

$$\text{True NO}_2 = \frac{1}{m_2} * \text{Aeroqual NO}_2 - \frac{m_3}{m_1 * m_2} * \text{Aeroqual O}_3 - \left( \frac{c_2}{m_2} - \frac{c_1 * m_3}{m_1 * m_2} \right) \quad (6)$$

where  $\frac{m_3}{m_1 * m_2}$  and  $\left( \frac{c_2}{m_2} - \frac{c_1 * m_3}{m_1 * m_2} \right)$  are constants.

Eq. (6) is similar to Eq. (2) produced in our previous method [16], except that there is now a coefficient for the Aeroqual NO<sub>2</sub> term that



**Table 2**

OLS linear regression parameters (and 95% confidence intervals), coefficient of determination and summary statistics for adjusted Aeroqual NO<sub>2</sub> and O<sub>3</sub> concentrations compared to reference analyser concentrations. The calibration adjustments applied were: the second 48 h of each deployment period corrected using the unique calibration derived using the first 48 h of that deployment period (*Aq\_corr\_u*); the second 48 h of each deployment period corrected using the calibration derived when the first 48 h of data from all studies were combined (*Aq\_corr\_a*); and the second 48 h of each deployment period corrected using the calibration derived from the first 48 h in the first (November) study only (*Aq\_corr\_N*).

Monitor	Correction	Slope [95 % C.I.]	Intercept [95 % C.I.] / $\mu\text{g m}^{-3}$	R <sup>2</sup>	RMSE / $\mu\text{g m}^{-3}$	MB / $\mu\text{g m}^{-3}$	NMB	n
O <sub>3</sub> 3	<i>O<sub>3</sub>corr_u</i>	1.07 [1.06, 1.09]	−5.75 [−6.77, −4.72]	0.98	4.59	−2.10	−0.04	286
	<i>O<sub>3</sub>corr_a</i>	0.97 [0.94, 1.00]	−0.89 [−2.50, 0.72]	0.94	6.25	−2.17	−0.04	286
	<i>O<sub>3</sub>corr_N</i>	1.06 [1.03, 1.09]	−11.44 [−13.19, −9.68]	0.94	10.72	−8.54	−0.17	286
O <sub>3</sub> 4	<i>O<sub>3</sub>corr_u</i>	0.93 [0.91, 0.95]	−2.14 [−3.20, −1.07]	0.97	6.94	−5.56	−0.11	286
	<i>O<sub>3</sub>corr_a</i>	0.91 [0.89, 0.93]	−1.46 [−2.57, −0.35]	0.96	7.22	−5.68	−0.12	286
	<i>O<sub>3</sub>corr_N</i>	0.96 [0.94, 0.99]	−9.24 [−10.41, −8.07]	0.96	11.80	−10.99	−0.23	286
NO <sub>2</sub> 1	<i>NO<sub>2</sub>corr_u</i>	1.09 [0.98, 1.20]	1.25 [−2.06, 4.56]	0.59	15.75	3.65	0.14	286
	<i>NO<sub>2</sub>corr_a</i>	1.03 [0.98, 1.08]	3.60 [2.03, 5.17]	0.85	8.50	4.45	0.17	286
	<i>NO<sub>2</sub>corr_N</i>	1.33 [1.04, 1.63]	11.79 [2.62, 20.96]	0.22	47.23	20.47	0.78	286
NO <sub>2</sub> 2	<i>NO<sub>2</sub>corr_u</i>	0.80 [0.58, 1.02]	11.02 [4.29, 17.75]	0.16	31.69	5.79	0.22	286
	<i>NO<sub>2</sub>corr_a</i>	0.97 [0.89, 1.04]	11.46 [9.24, 13.68]	0.71	14.73	10.59	0.41	286
	<i>NO<sub>2</sub>corr_N</i>	0.87 [0.80, 0.93]	16.97 [15.09, 18.88]	0.73	16.21	13.47	0.52	286

may not be equal to 1. To account for this coefficient for Aeroqual NO<sub>2</sub> in practice, a multiple linear regression was fitted to the Aeroqual NO<sub>2</sub> values to derive three fitted parameters:

$$\text{Aeroqual NO}_2 = k_1 * \text{True NO}_2 + k_2 * \text{Aeroqual O}_3 + k_3 \quad (7)$$

The best estimates for  $k_1$ ,  $k_2$  and  $k_3$  were then used to calibrate Aeroqual NO<sub>2</sub> concentrations as follows:

$$\text{True NO}_2 = \frac{1}{k_1} (\text{Aeroqual NO}_2 - k_2 * \text{Aeroqual O}_3 - k_3) \quad (8)$$

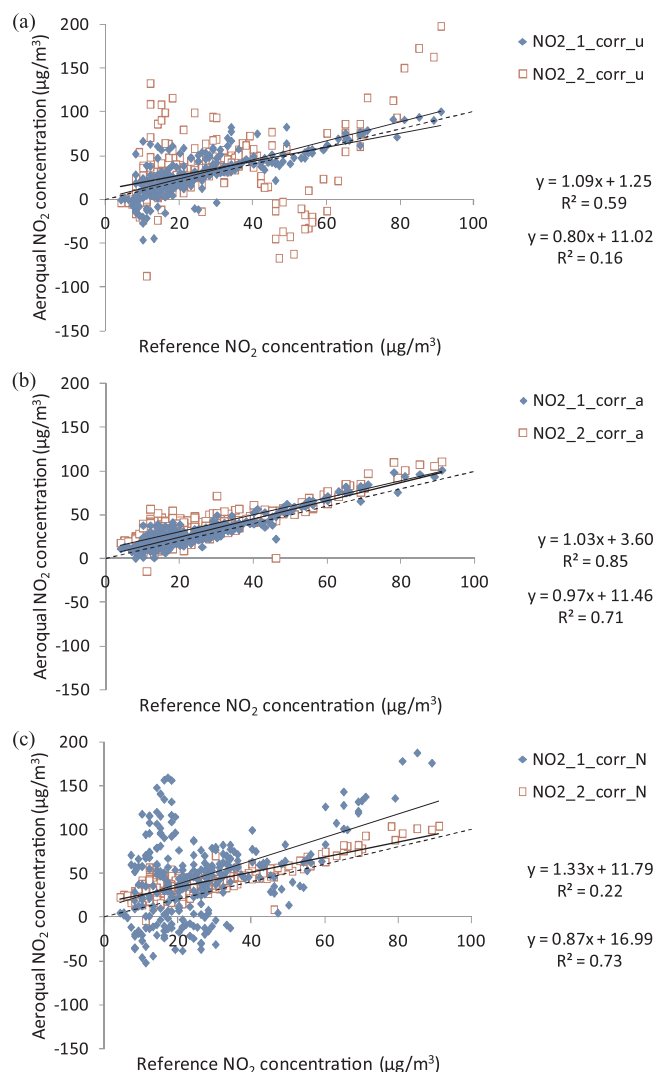
As for the O<sub>3</sub> monitors, the impact of temperature and relative humidity on NO<sub>2</sub> calibration regression equations was limited and inconsistent and was therefore not considered further.

The Aeroqual monitors were arbitrarily paired in this correction (reflecting the arbitrary pairing when a pair of monitors is used for field measurements), with NO<sub>2</sub>1 corrected using O<sub>3</sub>3 and NO<sub>2</sub>2 corrected using O<sub>3</sub>4. The three different calibration methods described in Section 3.1 were used; leading to Aeroqual NO<sub>2</sub> concentrations adjusted using the *Aq\_corr\_u*, *Aq\_corr\_a* and *Aq\_corr\_N* selections of training data (Table 1).

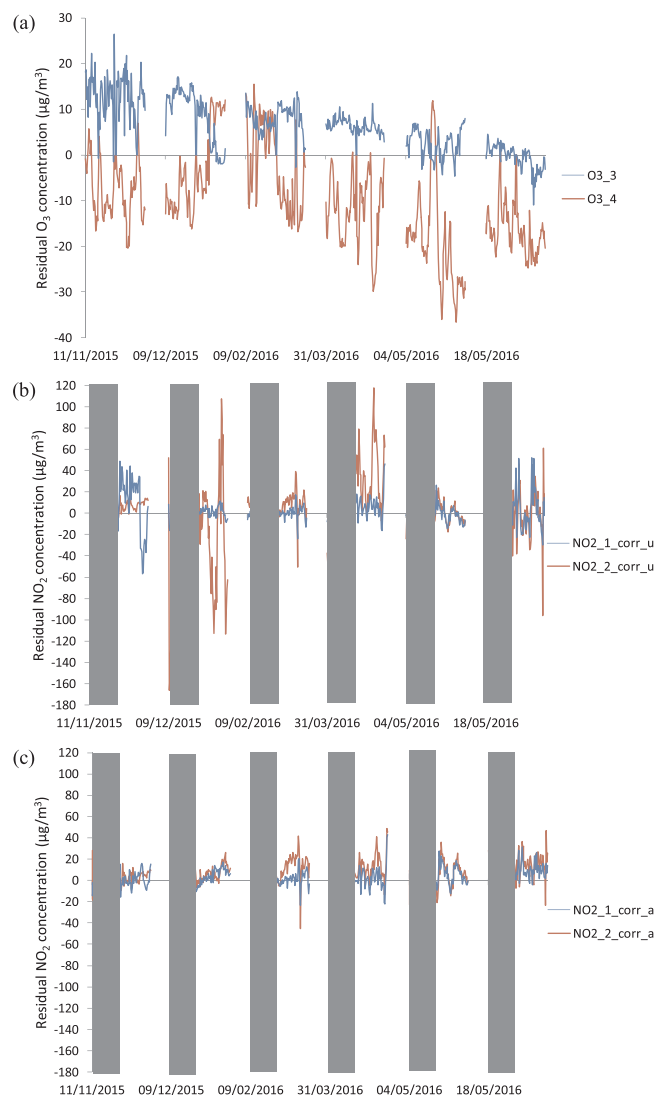
The adjusted Aeroqual NO<sub>2</sub> and Reference NO<sub>2</sub> concentrations were highly correlated over time, with the exception of some concentrations adjusted using the *Aq\_corr\_u* method (Fig. A5 in Supplementary material). The *Aq\_corr\_u* (NO<sub>2</sub>1 only), *Aq\_corr\_a* and *Aq\_corr\_N* (NO<sub>2</sub>2 only) calibration methods resulted in clearly defined linear relationships between adjusted Aeroqual and Reference NO<sub>2</sub> concentrations with R<sup>2</sup> values ranging from 0.70 to 0.85 (Fig. 3 & Table 2). The regression slope and intercept values were closest to 1 and 0, respectively, R<sup>2</sup> values highest, and RMSE values lowest for the *Aq\_corr\_a* selection of training data, for both Aeroqual NO<sub>2</sub> monitors (Figs. 3 & S7 and Table 2). The period of deviation between Aeroqual NO<sub>2</sub>2<sub>corr\_u</sub> and Reference NO<sub>2</sub> is from the December deployment, during which there was a very poor training data regression fit with small coefficients for *Reference NO<sub>2</sub>* and O<sub>3</sub>4 and a large offset (Table 1 & Fig. A6 in Supplementary material). The cause of the poor calibration regression during this particular deployment is not known but may be the result of transient errors in one or other of the Aeroqual monitor measurements in this period.

### 3.3. Temporal changes in Aeroqual O<sub>3</sub> and NO<sub>2</sub> monitor responses

We calculated the difference between unadjusted Aeroqual O<sub>3</sub> and Reference O<sub>3</sub> concentrations to assess drift in the output of the Aeroqual O<sub>3</sub> sensors during this study (Fig. 4a). The difference generally declined over the 6 months of measurements. This decline may have resulted



**Fig. 3.** Scatter plots of adjusted hourly Aeroqual NO<sub>2</sub> concentration estimates vs. reference analyser NO<sub>2</sub> concentrations for the test data (i.e. 2nd half of deployment periods only). (a) Aeroqual data adjusted using calibration derived from training data in each unique deployment period (*aq\_corr\_u*). (b) Aeroqual data adjusted using calibration derived from training data combined from all deployment periods (*aq\_corr\_a*). (c) Aeroqual data adjusted using calibration derived from the first deployment period (November) only (*aq\_corr\_N*).



**Fig. 4.** Residual plots of (a) unadjusted Aeroqual  $O_3$  – reference  $O_3$ ; (b)  $NO_2$   $aq\_corr\_u$  – reference  $NO_2$  and (c)  $NO_2$   $aq\_corr\_a$  – reference  $NO_2$  over the 96-h deployments. Both training and test data sets are included (training data indicated by grey boxes). Different deployment periods are separated by gaps in time series.

from deterioration in the sensitivity of the  $O_3$  monitors as the sensors aged. However, the calibration procedures overcame these effects to enable reasonably accurate representation of reference analyser concentrations.

We also plotted the difference between Aeroqual and Reference  $NO_2$  concentrations for Aeroqual concentrations adjusted using  $Aq\_corr\_u$  and  $Aq\_corr\_a$  training data selections (Fig. 4b and c). The residual concentrations varied more using the  $Aq\_corr\_u$  calibration method than for the  $Aq\_corr\_a$  method, with periods of large residual values corresponding to deployments with lower correlations between reference and Aeroqual  $NO_2$  concentrations in the data (Table 1). The residual  $NO_2$  concentrations showed no trend across the deployment periods for the  $Aq\_corr\_u$  correction method, as this calibration procedure partly corrected for the temporal drift in the Aeroqual  $O_3$  sensors identified above, while for the  $Aq\_corr\_a$  correction method the residuals generally increased within each individual deployment and across all the deployments.

Daily averages of the hourly residuals for the Aeroqual  $O_3$  and  $NO_2$  measurements adjusted using the first set of training data ( $Aq\_corr\_N$ ) were calculated for all of the available study data (including data for

deployments > 96 h in length) (Fig. A8 in Supplementary material). The daily  $O_3$  residuals became more negative during each deployment; however at the start of each deployment the residuals appeared to revert to close to zero, as measured for the first residual values in the  $Aq\_corr\_N$  calibration period in November 2015. This may be indicative of changes in the  $O_3$  sensor response after the monitor has been turned on, with repeated instances of this effect for individual deployments. The Aeroqual  $NO_2$  sensors exhibited a general increase in the daily residuals during each deployment that is likely to have resulted from the use of the Aeroqual  $O_3$  data in the calibration of the Aeroqual  $NO_2$  sensor.

Overall, these observations suggest that it is beneficial to make calibration and field measurements after consistent time periods following monitor start-up (as was done here).

#### 4. Discussion

As a result of technological developments portable monitors are increasingly used to measure ambient pollution concentrations over extended geographical areas, and for mobile and personal monitoring. Field calibration procedures are necessary to improve and quantify the accuracy of estimates of ambient pollution concentrations from portable monitors. In this study we investigated changes in calibration relationships for two pairs of Aeroqual  $O_3$  and  $NO_2$  monitors vs. reference analysers over a period of 6 months as the sensors approached the manufacturer's specified 2-year lifetime. We developed a multiple linear regression calibration method that allowed the Aeroqual  $NO_2$  sensor to have a regression slope with 'true'  $NO_2$  that is not equal to 1 (as was assumed in the method of Lin et al. [16]), due for example to sensor aging, in addition to allowance for a non-unity response to  $O_3$ . The use of multiple linear regression compared to linear regression to correct other portable  $NO_2$  sensors has been observed to produce higher correlation between calibrated sensor estimates and reference  $NO_2$  concentrations [15].

Changes in monitor responses over time (including apparent baseline drift in  $O_3$  sensor output, and discrepancies between the 2 Aeroqual  $NO_2$  sensors) resulted in relatively inaccurate concentrations estimates (cf. reference concentrations) from calibration equations derived in the first training period and applied to subsequent *test* deployments (e.g.  $NO_2$  RMSE =  $47.2 \mu\text{g m}^{-3}$  ( $n = 286$ ) for a dataset of all *test* periods combined, for one of the two monitor pairs). Substantial improvements in accuracy of estimated concentrations were achieved by combination of repeated intermittent training data into a single calibration dataset ( $NO_2$  RMSE =  $8.5 \mu\text{g m}^{-3}$  for same *test* dataset described above). This latter approach to field calibration is recommended. Our results also highlight the importance, when possible, of calibration of monitors in similar urban pollution environments to where they are to be used, as has been adopted in previous studies [9,10,16]. One recent study [20] used un-calibrated concentrations from the Aeroqual monitors to measure  $NO_2$  and  $O_3$  in indoor office environments with low  $O_3$  concentrations which may have removed the need to adjust the Aeroqual  $NO_2$  monitors for  $O_3$  cross-sensitivity.

Our observations suggest that a single calibration study may not be able accurately to represent the relationship between monitor and reference analyser over an extended period, for example because of unidentified measurement problems during a particular calibration deployment period. In contrast, we have observed that combination of several short calibration deployments over a period of 6 months into a single calibration dataset provided adjusted Aeroqual concentration estimates that agreed more closely with the reference analyser concentrations, provided that for the  $NO_2$  monitors concurrent  $O_3$  monitor measurements were also used as input to the calibration relationship. We observed differences in Aeroqual  $NO_2$  calibrations over shorter time periods (e.g. within 1 month), which may be attributable to differences in the relative pollutant concentration ranges during the calibration periods (it is important to recognise the importance of pollutant

concentration ranges when assessing percentages of explained variation in pollution concentrations with  $R^2$  statistics, and in relation to potential inaccuracies when extrapolating outside of calibration concentration ranges). During a winter pollution episode with elevated  $\text{NO}_2$  concentrations none of the calibration methods tested for the  $\text{O}_3$  monitor could accurately represent the analyser concentrations, again highlighting the importance of (when possible) calibrating the Aeroqual monitors using pollution concentrations similar to those anticipated during intended field application.

It is important to acknowledge: (a) that sensor drift is likely to exist; (b) it may well not be possible to predict the nature of the sensor drift *i.e.* the rate of sensor drift may not be constant (this is evident in the irregular nature of the drift observed in our measurements (Fig. 4a)); so (c) for practical use in the field these sensors must be (re)calibrated against reference analysers as often as possible. Further refinements to the relatively simple and parsimonious approach we have taken to sensor calibration may be possible, including attempts to characterise the irregular temporal drift in the  $\text{O}_3$  sensors, and examination of the effect of reduced frequency calibration on the accuracy of sensor estimates.

In conclusion, our observations indicate that the benefit [in terms of improved accuracy of calibrated  $\text{NO}_2$  estimates] of calibration equations averaged by pooling data from a number of calibration periods over a period of a few months (increased size of calibration dataset and potentially increased concentration range) outweighed the disadvantage arising from any monitor drift throughout the overall measurement period. It is possible that the combination of several calibration periods interspersed within the overall measurement period mitigated the risk of only relying on calibration data from short periods when the range of concentrations may have been limited or otherwise unrepresentative. Therefore, out of the 3 calibration methods for  $\text{NO}_2$  sensors evaluated in this study, pooling of interspersed calibration data is recommended as a pragmatic approach to calibration of Aeroqual monitors for field deployment. Although the sensors used in this research were approaching the 2-year working lifetime specified by the manufacturer (Aeroqual, 2015 personal communication), our observations suggest that frequent intermittent calibrations as presented above may allow the sensors to remain useable beyond this specified lifetime.

## Acknowledgements

Nicola Masey is funded through a UK Natural Environment Research Council CASE PhD studentship (NE/K007319/1), with support from Ricardo Energy and Environment. Jonathan Gillespie is funded through an Engineering and Physical Sciences Research Council Doctoral Training Grant EPSRC DTG (EP/L505080/1 and EP/K503174/1) studentship, with support from the University of Strathclyde and Ricardo Energy and Environment. Eliani Ezani is funded by the Ministry of Higher Education Malaysia (KPT(BS)860126295394). Chun Lin is funded through NERC/Innovate UK grant NE/N007352/1. Hao Wu is funded by the University of Edinburgh and the NERC Centre for Ecology & Hydrology (NERC CEH project number NEC04544). We acknowledge access to the AURN measurement data, which were obtained from [uk-air.defra.gov.uk](http://uk-air.defra.gov.uk) and are subject to Crown 2014 copyright, Defra, licensed under the Open Government Licence (OGL). The research data associated with this paper are available at: <https://doi.org/10.15129/a5820b14-367f-47f3-b0b9-beabcb347355>.

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.snb.2018.07.087>.

## References

- [1] J. Gillespie, N. Masey, M.R. Heal, S. Hamilton, I.J. Beverland, Estimation of spatial patterns of urban air pollution over a 4-week period from repeated 5-min measurements, *Atmos. Environ.* 150 (2017) 295–302, <https://doi.org/10.1016/j.atmosenv.2016.11.035>.
- [2] J. Gillespie, I.J. Beverland, S. Hamilton, S. Padmanabhan, Development, evaluation, and comparison of land use regression modeling methods to estimate residential exposure to nitrogen dioxide in a cohort study, *Environ. Sci. Technol.* 50 (2016) 11085–11093, <https://doi.org/10.1021/acs.est.6b02089>.
- [3] N. Masey, J. Gillespie, M.R. Heal, S. Hamilton, I.J. Beverland, Influence of wind-speed on short duration  $\text{NO}_2$  measurements using Palmes and Ogawa passive diffusion samplers, *Atmos. Environ.* (2017), <https://doi.org/10.1016/j.atmosenv.2017.04.008>.
- [4] M.R. Heal, M.A. O'Donoghue, J.N. Cape, Overestimation of urban nitrogen dioxide by passive diffusion tubes: a comparative exposure and model study, *Atmos. Environ.* 33 (1999) 513–524, [https://doi.org/10.1016/S1352-2310\(98\)00290-8](https://doi.org/10.1016/S1352-2310(98)00290-8).
- [5] E.G. Snyder, T.H. Watkins, P.A. Solomon, E.D. Thoma, R.W. Williams, G.S.W. Hagler, D. Shelov, D.A. Hindin, V.J. Kilaru, P.W. Preuss, The changing paradigm of air pollution monitoring, *Environ. Sci. Technol.* 47 (2013) 11369–11377, <https://doi.org/10.1021/es4022602>.
- [6] M. Bart, D.E. Williams, B. Ainslie, I. McKendry, J. Salmond, S.K. Grange, M. Alavi-Shostari, D. Steyn, G.S. Henshaw, High density ozone monitoring using gas sensitive semi-conductor sensors in the Lower Fraser Valley, British Columbia, *Environ. Sci. Technol.* 48 (2014) 3970–3977, <https://doi.org/10.1021/es404610t>.
- [7] I. Heimann, V.B. Bright, M.W. McLeod, M.I. Mead, O.A.M. Popoola, G.B. Stewart, R.L. Jones, Source attribution of air pollution by spatial scale separation using high spatial density networks of low cost air quality sensors, *Atmos. Environ.* 113 (2015) 10–19, <https://doi.org/10.1016/j.atmosenv.2015.04.057>.
- [8] M.I. Mead, O.A.M. Popoola, G.B. Stewart, P. Landshoff, M. Calleja, M. Hayes, J.J. Baldovi, M.W. McLeod, T.F. Hodgson, J. Dicks, A. Lewis, J. Cohen, R. Baron, J.R. Saffell, R.L. Jones, The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks, *Atmos. Environ.* 70 (2013) 186–203, <https://doi.org/10.1016/j.atmosenv.2012.11.060>.
- [9] A. Wang, M. Fallah-Shorshani, J. Xu, M. Hatzopoulou, Characterizing near-road air pollution using local-scale emission and dispersion models and validation against in-situ measurements, *Atmos. Environ.* 142 (2016) 452–464, <https://doi.org/10.1016/j.atmosenv.2016.08.020>.
- [10] L. Deville Cavellin, S. Weichenthal, R. Tack, M.S. Ragettli, A. Smargiassi, M. Hatzopoulou, Investigating the use of portable air pollution sensors to capture the spatial variability of traffic-related air pollution, *Environ. Sci. Technol.* 50 (2016) 313–320, <https://doi.org/10.1021/acs.est.5b04235>.
- [11] A. Lewis, P. Edwards, Validate personal air-pollution sensors, *Nature* 535 (2016) 29–31, <https://doi.org/10.1038/535029a>.
- [12] A.C. Lewis, J.D. Lee, P.M. Edwards, M.D. Shaw, M.J. Evans, S.J. Moller, K.R. Smith, J.W. Buckley, M. Ellis, S.R. Gillot, A. White, Evaluating the performance of low cost chemical sensors for air pollution research, *Faraday Discuss.* 189 (2016) 85–103, <https://doi.org/10.1039/C5FD000201J>.
- [13] N. Masson, R. Piedrahita, M. Hannigan, Quantification method for electrolytic sensors in long-term monitoring of ambient air quality, *Sensors* 15 (2015) 27283–27302, <https://doi.org/10.3390/s151027283>.
- [14] X. Pang, M.D. Shaw, A.C. Lewis, L.J. Carpenter, T. Batchellier, Electrochemical ozone sensors: a miniaturised alternative for ozone measurements in laboratory experiments and air-quality monitoring, *Sens. Actuators B Chem.* 240 (2017) 829–837, <https://doi.org/10.1016/j.snb.2016.09.020>.
- [15] L. Spinnelle, M. Gerboles, M.G. Villani, M. Aleixandre, F. Bonavitacola, Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: ozone and nitrogen dioxide, *Sens. Actuators B Chem.* 215 (2015) 249–257, <https://doi.org/10.1016/j.snb.2015.03.031>.
- [16] C. Lin, J. Gillespie, M.D. Schuder, W. Duberstein, I.J. Beverland, M.R. Heal, Evaluation and calibration of Aeroqual series 500 portable gas sensors for accurate measurement of ambient ozone and nitrogen dioxide, *Atmos. Environ.* 100 (2015) 111–116.
- [17] R.M. Duvall, R.W. Long, M.R. Beaver, K.G. Kronmiller, M.L. Wheeler, J.J. Szykman, Performance evaluation and community application of low-cost sensors for ozone and nitrogen dioxide, *Sensors* 16 (2016) 1698, <https://doi.org/10.3390/s16101698>.
- [18] W. Jiao, G. Hagler, R. Williams, R. Sharpe, R. Brown, D. Garver, R. Judge, M. Caudill, J. Rickard, M. Davis, L. Weinstock, S. Zimmer-Dauphinee, K. Buckley, Community Air Sensor Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern United States, *Atmos. Meas. Technol.* 9 (2016) 5281–5292, <https://doi.org/10.5194/amt-9-5281-2016>.
- [19] DEFRA, The Data Verification and Ratification Process, (2017) (Accessed 1 April 2017), [https://uk-air.defra.gov.uk/assets/documents/The\\_Data\\_Verification\\_and\\_Ratification\\_Process.pdf](https://uk-air.defra.gov.uk/assets/documents/The_Data_Verification_and_Ratification_Process.pdf).
- [20] T. Sziget, C. Dunster, A. Cattaneo, A. Spinazzè, C. Mandin, E. Le Ponner, E. de Oliveira Fernandes, G. Ventura, D.E. Saraga, I.A. Sakellaris, Y. de Kluienaar, E. Cornelissen, J.G. Bartzis, F.J. Kelly, Spatial and temporal variation of particulate matter characteristics within office buildings—the OFFICAIR study, *Sci. Total Environ.* 587–588 (2017) 59–67, <https://doi.org/10.1016/j.scitotenv.2017.01.013>.